AMENDMENTS TO THE SPECIFICATION

1. Please replace paragraph [0020] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

> [0020] As is well known to those skilled in the art, voiced sounds such as speech are produced by an oscillation of the vocal cords that modulates airflow into quasi-periodic pulses which excite resonances in the vocal tract. The rate of these pulses is generally called the fundamental frequency or "pitch." In general, the periodicity, or "pitch period" of a voiced audio signal represents the time between the largest magnitude positive or negative peaks in a time domain representation of the voiced audio signal. Although speech signals are not actually perfectly periodic, the estimated pitch frequency and its reciprocal, the pitch period, are still very useful in modeling the speech signal. Note that the reminder of the discussion makes reference to both pitch and pitch period. There are highly elaborate methods for determining pitch; however, as these concepts are well known to those skilled the art, the determination of pitch and pitch period described herein will be a basic one, based simply on finding the peak of cross correlation. However, it should be clear in view of the discussion provided herein that that any conventional method for determining pitch and pitch period may be used <u>in</u> the temporal audio scaler.

2. Please replace paragraph [0052] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

> [0052] In addition, the computer 110 may also include a speech input device, such as a microphone 198 or a microphone array, as well as ~~and~~ a loudspeaker 197 or other sound output device connected via an audio interface 199. Other input devices (not shown) may include a joystick, game pad, satellite dish, scanner, radio receiver, and a television or broadcast video receiver, or the like. These and other input devices are often connected to the processing unit 120 through a user input interface 160 that is coupled to the system bus 121, but may be connected by other

interface and bus structures, such as, for example, a parallel port, game port, or a universal serial bus (USB). A monitor 191 or other type of display device is also connected to the system bus 121 via an interface, such as a video interface 190. In addition to the monitor, computers may also include other peripheral output devices such as a printer 196, which may be connected through an output peripheral interface 195.

3. Please replace paragraph [0057] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0057] The more traditional application of time-scale modification of audio signals is in slowing down or speeding up the overall time scale of a signal, many times to reduce listening time, or to improve intelligibility. Besides that application, in the last few years time-scale modification of audio signals containing speech has also been used for improving the quality of signals transmitted across lossy and delay prone packet-based networks such as the Internet and then reconstructed on a client computer or receiver. For example, in many applications it is desirable to stretch or compress one or more frames of an audio signal containing speech.

4. Please replace paragraph [0118] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0118] Further, unlike the windowing used for voiced segments, a preferred overlapping smoothing window used is different here. For example, while the overlapping portions of the signal used for stretching the voiced segments are correlated, the overlapping portions of the signal in the ~~invoiced~~ unvoiced case are theoretically uncorrelated. Therefore, better results, i.e., reduced artifacts, are achieved at boundary points by using a window such as a conventional sine ~~widow~~ window which keeps the energy constant and sums to one when squared and added, i.e., $(wa[n])^2+(wb[n])^2=1$. Such windows are well known to those skilled in the art. This process is generally represented by steps 400 through 480 of FIG. 4.

5. Please replace paragraph [0120] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0120] Next, given x[n], whether or not it has been zero padded 410, the FFT is computed 420. The phase of this FFT is then randomized 430. Next, the inverse FFT, y[n], is computed 440 from this FFT having the randomized phase. The result of this process, steps 420 through 440, is a synthetic frame or segment, y[n], having a similar spectrum, but no correlation with the original segment, x[n]. The original (non-zero padded) frame or segment x[n] is then split into two parts, and y[n] is inserted between those two parts, and seamlessly added using the aforementioned conventional overlap/add process 450, such as, for example, a conventional sine ~~widow~~ window to create a stretched frame.

6. Please replace paragraph [0129] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0129] At this point, the phase of the resulting FFT, $Z[w]$, is then randomized 530, scaled to compensate for the smoothing window gain (e.g., 2 for a sine window), and the inverse FFT, $u[n]$, is computed 535 from $Z[w]$ to create a synthetic sub-segment having a similar spectrum, but no correlation with the original segment, $z[n]$. The newly synthesized signal sub- segment, $u[n]$, is then inserted into the original signal at position s, and seamlessly added using the aforementioned conventional overlap/add process 540, such as, for example, a conventional sine ~~widow~~ window to create a partially stretched frame, as illustrated by Equation 7:

$$y[(i \times k+1):(i \times k+2K)] = y[(i \times k+1):(i \times k+2K)] + w[1:2K] \times u[1:2k] \quad \text{Equation 7}$$

7. Please replace paragraph [0137] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0137] At this point, similar to what was described above, the phase of the resulting FFT, $Z[w]$, is then randomized 640, and the inverse FFT, $u[n]$, is computed 645 from $Z[w]$ to create a synthetic sub-segment having a similar spectrum, but no correlation with the original segment, $z[n]$. The newly synthesized signal sub- segment, $u[n]$, is then inserted into the original signal at position s, and seamlessly added using the aforementioned conventional overlap/add process 650, such as, for example, a conventional sine ~~widow~~ window to create a partially stretched frame, as illustrated by Equation 7, as described above.

8. Please replace paragraph [0145] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0145] In particular, as illustrated by FIG. 7, in selecting 710 the best points for stretching the current signal frame, the process begins by determining 700 a total number of internal segments T in a desired frame size, M, $(T=(M/K)-1))$, and a total number of internal segments P in the original frame size, I , $(P=(M/K)-1))$. At this time, a point counter Pt is set to $P+1$ 720. Next, the average energy $E(i)$ of each sub-segment is computed 730 as illustrated by Equation 10:

$$E(i) = avg(x(s[i]:s[i+1])^2) \qquad \text{Equation 10}$$

9. Please replace paragraph [0148] of United States Patent Application Publication No. 20050055204 with the following amended paragraph:

[0148] Once weighted 740, the average energy values $E(i)$ are examined to select a segment $s[j]$ having the lowest energy value 750. As noted above, these lowest energy segments are then split 750 into two, with a new starting point $s[Pt]$ for

stretching the current frame being located at the split point as illustrated by Equation 12:

$$s[Pt] = (s[j] + s[j+1])/2 \qquad \text{Equation 12}$$